This work is licensed under a
Creative Commons Attribution-NonCommercial-
NoDerivs 3.0 Licence.

To view a copy of the licence please see:
http://creativecommons.Org/licenses/by-nc-nd/3.0/

(832)

INSTITUTE FOR DEVELOPMENT STUDIES
UNIVERSITY OF NAIROBI

IDS LIBRARY
RESERVE COLLECTION

Discussion Paper No. 106

Cross-Tabulation of Means: Description of Computer Program 'S04B'.

BY

Donald Shepard
Ken Prewitt

May, 1971

Any views expressed in this paper are those of the authors. They should not be interpreted as reflecting the views of the Institute for Development Studies or of the University of Nairobi.

Note: This program description is designed to instruct new new comers to the computer, or to this program, and to provide reference information and hints to experienced users and programmers. The experienced user may find it sufficient to skim the expository section of the paper, and turn to the appendices at the end.

Contents:

Introduction.

Let us assume that an investigator wishes to understand if student scores on the H. S. C. examination are related to family background, especially to whether the student comes from an educated family. He selects a sample of 50 students from a Nairobi school and obtains the following information about each student: The level of schooling his father reached.
The level of schooling his mother reached.
How many brothers or sisters have completed primary school.
Sex of the student.
Examination score of the student.

The computer program S04B allows the investigator to determine whether student performance on the examination can be explained with reference to the type of background factors listed above. Two types of analysis are possible.

First, it is possible to construct a simple frequency table which would show what percentage of students with, for instance, educated fathers scored high on the examination compared to the percentage of students with uneducated fathers. We call this procedure simple table analysis, and discuss it below in connection with the machine instructions which will produce simple frequency tables.

Second, using a more powerful analysis possible with the program, the investigator can determine the mean score of a group of students identified in terms of two independent variables. Let us give an example. Presume that the education of the parents has been coded as follows: No education, Some education though not beyond primary six, Education beyond primary six. If the education of the father and mother were cross-tablulated, nine cells would be produced, as indicated:

		Father's Education		
		None	Prim. Only	More than Prim.
Mother's Education.	None	1	2	3
	Prim. Only	4	5	6
	More th. Prim.	7	8	9

S04B would print the mean (average) examination score for all the students which fall into cell 1 (None, None), all that fall into cell 2 (Mother None, Father Primary), and so forth. The investigator would then learn whether the average examination score of students from different types of families differed in any significant manner. For instance, if the average examination score of students in cell 9 (both parents educated) is much higher than the average score of students in cell 1 (neither parent educated), he might conclude that family background does indeed affect examination performance. (In addition to the means, the program prints a table of cell variances; a technical discussion of variance cannot be included in this description of the program.)

The Data.

The information which is collected about each of the fifty students must be transferred to computer cards (or tape) of course. For the reader unfamiliary with computer cards, we here include a brief description of how the data might be recorded. A computer card has 80 columns, each of which has a place to punch a 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, or may be left blank. Although alphabetic data could be entered on cards as well, it cannot be read by this program, and will cause the computer to stop. A codebook for the study used to illustrate this program might be set up as follows:

Cols. 1 through 2 -- identification number for each student (from 01 to 50).

Col.4 -- Father's Education.
punch 0-- no education
1-- some education, though not beyond primary six.
2-- education beyond primary six
9-- no information on father's education

Col.5 -- Mother's Education.
coded in same way as father's education

Col. 6-- No. of brothers & sisters who have completed primary school.

0 -- none
1 -- one
2 -- two
3 -- three
4 -- four
5 -- five or more
9 -- no information.

Col. 7 -- Sex of the student
0 -- male
1 -- female
9 -- no information

Col. 9-10 -- Examination score.

The actual score is recorded; the lowest score is a 6 and the highest score is a 54. If there is "no information" on a student's examination score, the special code 99 is entered in columns 9-10. Otherwise, every student will have some number between 6 and 54 punched into columns 9 and 10.

A data card might look as follows:

```
Column  1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 .....
        2 3   2 1 0 2   3 4
```

The first two columns indicate that this is student no. 23 in the sample (the identification number), the next column is blank, the 2 in col. four indicates the education level of the father, the 1 in col. five indicates education level of the mother, the 0 in col.6 indicates that no brothers or sisters have completed primary schooling, the 1 in col. 7 indicates that the student is a female, col. 8 is blank, and cols. 9 & 10 indicate that the examination score is 34.

Control Cards.

A total of nine different control cards (or types of control cards) must be prepared to execute the program correctly. All numbers entered onto the control cards must be right-justified (placed as far to right of the appropriate coding field as possible). In general, five columns are allowed for numbers. Numbers may not be punched with a decimal point except on card 6 (Category Maxima), where the decimal point is optional and ten columns are allowed for each number.

Cols. 73-80 of all control cards are reserved for sequence numbers. The first control card should be labeled 010 in cols. 78-80, the second 020, the third 030, etc. This allows control cards to be easily identified.

Card 1 - The Title Card.

This easily-prepared card identifies the name of the job (e.g., S05D), and describes the job in such as way as to allow the investigator to identify and perhaps file his print-out.

The title card also includes the name of the program, S04B in the present case. Set-up is as follows:

Cols.	Description.
1-4	Name of job
6-72	Description of the job; used to facilitate subsequent identification of print-out. You may wish to include, for instance, your name, the project you are working on, what type of analysis this run is, and the date.
78-80	Sequence number, i.e., 010.

Card 2 - The Parameters Card.

This card reports basic information about the data. Set-up is as follows:

Cols.	Parameter	Description
1-5	CASES	The total number of cases. In our example, 50 students are being studied and thus the number 50 will appear in columns 4-5 (because the number must be right justified). The number of cases must not exceed 99999.
8-10	VARS	The number of variables (or separate pieces of information) recorded for each case. In the example, there are five variables: father's education, mother's education, number of siblings with education, sex, examination score. The number 5, then would appear in col. 10. The number of variables must not exceed 100.
14-15	DEPS	The number of dependent variable cards. The term dependent variable is used in this program to refer to variables on which means are computed, similar to what was suggested in the introductory section. The meaning of dependent variable will become clearer in the discussion of the "dependent variable card." Because in our example, the examination score of the students is the only dependent variable, the number 1 would be entered in col. 15. The number of dependent variable cards must be at least 1, and must not exceed 20.
20	UNIT	Input unit for the data. This parameter tells the computer whether your data are recorded on computer cards, magnetic tape, or paper tape. Use the following code: 1=cards; 2=magnetic tape; 4=paper tape.
24-25	RCOD	The number of variables to be created by recording the values of original variables. The recode procedure will become clearer when we discuss cards 5 and 6. The number of recoded variables must not exceed 40, and is 0 if no recoded variables are to be created.

Cols.	Parameter	Description (cont.)
30	LIST	Sometimes it is valuable for the investigator to list all of his data cards, especially if he has reason to think that they may contain errors. The program has the option of listing data cards. He may list all of them or only those on which some variable exceeds the maximum value allowed. (The maximum value is described in connection with card 3.) 0=no list; 1=list all cases for which any variable exceeds its specified maximum variable; 2=list all cases.
35	REW	If input data are in the form of magnetic tape, it may be necessary to rewind the tape if more than one set of analyses is to be performed. If you request the tape to be rewound, this would happen after card 8. You would then have a second set of control cards, just as the first set, which would describe the analysis to be accomplished on the second run. 0= no rewind; 1= yes, rewind.
38-40	TVAR	This parameter identifies the total number of variables, which is the sum of the original input variables plus the recoded variables. In our example, we will recode one variable, thus the number to appear in col.40 will be a 6, the five original variables plus the one new variable. The total number of variables must not exceed 100. (The number of original variables also must not exceed 100. Thus if your initial data file included 95 variables and you wished to create 10 new, recoded variables, you would be exceeding the limitations of the program. It would be necessary to reduce your initial file to only 90 variables, which you could easily do by declaring certain variables, that is columns, to be skipped on the Format Card, described below. The variables skipped would of course be variables in which you had no interest for this particular analysis.)
49-60	TAPE	Name of the magnetic tape with your data. Blank if the unit input is not a 2.
78-80		Sequence number, i.e., 020.

Card 3 -- The Format Cards

The computer must know in what columns your variables appear. This requires a format card, which is described in more detail in Appendix 2. In this program, two format cards are always required. The format statement appears in cols. 1-72 of the first card and continues, if necessary, in cols. 1-72 of the

second card, treating the first col. of the second card just as if it were col. 73 of the initial card. If your format statement requires only one card, then insert a card blank in columns 1 to 72 as the second one.

The number of "F" fields identified in the format card must agree with the number of input variables (VARS) identified on the parameter card (not, of course, with the TVAR on the parameter card.) In the example used above, the Format card would appear as follows:

(3X,4F1.0,1X,F2.0)

Because the statement is easily put on one card, it would be followed by a blank card. The format statement says that the first three columns are to be skipped. (Actually the first two columns include an identification number, but you are uninterested in the identification of the student for analysis purposes and thus you simply skip it.) The next four columns are four separate variables, the next column is skipped, and the next two columns are to be read as a single variable. This format statement tells the program that col. 4 is the first variable (father's education), col.5 is the second variable (mother's education), and so forth, with cols. 9-10 being the fifth variable (exam. score).

The set-up of card 3 is as follows:

Columns	Description
(First Format card)	
1 - 72	Format statement
78 - 80	Sequence number, i.e. 030
(Second Format card)	
1 - 72	Continuation of Format statement if necessary. Otherwise blank.
78 - 80	Sequence number, i.e. 040.

Card 4 -- The maximum Values Card

It is necessary to instruct the program regarding the maximum value which each of the variables takes. The maximum value of the original input variables as well as any recoded variables must be included. Thus the number of maximum values identified will correspond with the TVAR (total variables) noted on the parameter card. The maximum value listed for a particular variable need not be the true maximum value of that variable; for

instance, if you wished to eliminate from analysis any cases in which father's education was not known, the maximum value of the first variable would be 3 rather than 9. Cases in which 9 was punched in col. 4 (indicating that father's education was not known) would not appear in the tables, because the 9 punch of course exceeds what has been set as the maximum value.

If a variable is not being used as an independent variable and its maximum is not known (or higher than 99999), the number -1 may be entered as the maximum value. The same designation (-1) may be used to by-pass a particular variable if you select to LIST cases in which variables exceed their maximum values. The variable coded -1 would not cause a case to be listed. (All of this implies that you have selected "1" as your list option on the parameter card.)

The maximum value of each variable is entered in a five-colum field, the maximum value of the first variable appearing in cols. 1-5, the maximum value of the second variable appearing in cols. 6-10, etc. You may use only columns 1 through 70, thus you can record 14 maximum values on one card. If you have more than 14 variables, simply continue in col. 1-5, etc. of the next card.

The maximum values are entered in sequence order, beginning with all of the original input variables followed by the recoded variables. The set-up of card 4 is as follows:

Cols.	Parameter	Description
1-5	MAX VALUES	The maximum value of the first variable.
6-10		The maximum value of the second variable.
.		
.		
.		
66-70		The maximum value of the fourteenth variable.
78-80		Sequence number.

In the example we have been using, the maximum value card might look as follows:

. . . . 2 . . . . 2 . . . . 5 . . . . 9 . . . 54 2

The maximum values of father's and mother's education are set at 2, thus eliminating the cases with no information; the maximum value of the next variable (number of brothers and sisters educated) is set at 5, again eliminating cases with no

information; the maximum value of the variable sex is set at 9, this time including cases with no information; the maximum value of the fifth variable, examination score, is 54 because that is the highest valid score. Cases with a score above 54, i.e. those coded 99, will be excluded from analysis*. The maximum value of the next variable, a recoded variable, is 2, the reason for this will become clear when we discuss recode procedures.

In order to minimise errors in establishing the maximum values (and errors in identifying any parameters of your data) it is necessary to prepare a sheet which identifies all basic information ablut the data. Such a sheet might take the following form:

Name of Variable	Sequence Number	Cols.	Coding Range	Maximum Value for Analysis
Father's Education	1	4	0 - 9	2
Mother's Education	2	5	0 - 9	2
Educated Siblings	3	6	0 - 9	5
Student's Sex	4	7	0 - 9	9
Examination Score	5	9-10	6 - 99	54
1st Recode: Collapse into three categories the var. on educated siblings	6	-	0 - 2	2

Card 5 -- The Recoded Variables Card

Often the investigator wishes to reduce the number of categories which appear in one of his variables. This is especially important in constructing tables, for it is often necessary to reduce the number of cells so that sufficient number of cases appear in each cell to make analysis possible. Consider two variables we have introduced: the education of the father and the number of siblings who have completed primary schooling. A table which used all of the values in these two variables would produce 18 cells (3 X 6), far too many for a study which has only 50 respondents.

Let us assume, then, that you wish to recode the variable "number of brothers and sisters completing primary school" into only three categories: one category of all those students who have none or only one sibling who has completed primary (values 0 or 1 for variable 3); one category for students with two or

* The maximum value for dependent variables should always exclude the code for "don't know" or "no information". Otherwise code 99 would be treated as a valid examinatiom score, and the averages would be meaningless.

three educated siblings (punches 2 and 3 in variable 3); and one category for students with four or more (punches 4 and 5). The recode card allows you to collapse your variable in this manner. Each variable to be recoded requires one card 5, on which is listed the variable to be recoded and the number of categories into which the original variable is to be recoded, followed by one or more card 6s, which instruct the machine just how to collapse the original variable.

On card 5, place the sequence number of the original variable in cols 3-5. In our example, we wish to recode the variable on the number of siblings with education, which is variable no. 3. (It can be seen that the basic information sheet prepared for your entire data set is useful for knowing the sequence number of each of your variables. This is especially important if your study includes more variables than just a few, which is nearly always the case.) In cols. 8-10 of card 5, place the number of categories to appear in the recoded variable. In our example, we are collapsing the original variable into three categories and thus the number 3 would appear in col. 10.

The card set-up for card 5 is as follows:

Cols.	Parameter	Description
3-5	ORIG	The sequence number of the original input variable which is to be recoded.
8-10	CATS	The number of categories created on the recoded variable. (Each category uses one cell of storage, see Appendix 5)
78-80		Sequence number.

Card 6 -- Category Maxima Card

Card 5 will be followed by one or more card sixes, which provide instructions regarding the way in which the original variable is to be collapsed. We have suggested that the original input variable 3 is to be collapsed as follows: punches 0 and 1 into one category (few educated siblings), punches 2 and 3 into another category (some educated siblings), punches 4 and 5 into another category (many educated siblings). On card six, the first 10 columns give the highest value from the original variable which is to go into the first category of the recoded variable. In our example, a 1 would appear in cols. 1-10 because a

score 1 is the highest score we allow any student to have who is to appear in the initial category of the new variable (the category of few educated siblings). The next 10 columns, 11-20, are used to record the highest value from the original variable which is to go into the second category of the new variable. In our example, the number 3 would appear in these columns. The next 10 columns, 21-30, are used to record the highest value from the original variable which is to go into the third category of the new variable. It can be seen that the total number of ten-column fields used will match the number of CATS identified in card 5.

Only 7 ten-column fields can be used on any card 6. If you wish to recode a variable into more than 7 new categories, use a second card 6; to recode into more than 14 new categories, use a third card 6, etc.

The actual values which are stored in the machine for recoded variables begin with 0 and proceed sequentially. Thus your recoded variable will be stored as follows:

Recoded Variable

0 = all cases punched 0 or 1 in variable 3*
1 = all cases punched 2 or 3 in variable 3
2 = all cases punched 4 or 5 in variable 3

This is important to keep in mind. We saw in connection with card 4 that every variable, including recoded variables, must be given a maximum value. The maximum value of this recoded variable, then, is a 2, because that is the highest score any respondent will take following the recoding procedures.

The program also has the facility for creating an "open-ended" category, which includes all cases with all the original variable above the last specified category maximum. This open-ended category is denoted by the next sequential value of the recoded variable. In the example above, the new category created would be:

Recoded Variable :

3 = All cases punched 6 or above (i.e. punched 9) in variable 3

To use this facility, the maximum value of the recoded variable

*Negative values of variable 3, if they existed would be placed in the first category of the recoded variable.

would be entered as 3 on card 4. Note that when this facility is used, the maximum value of a recoded variable on card 4 is the same as the value of CATS on card 5. (Otherwise, the maximum value on card 4 would be equal to CATS - 1.)

Note that in this example, as for any variable for which "no information" is coded as a high value, the "open-ended" case will be the "no information" cases. It is often better to create a separate category in this manner instead of excluding them entirely from a table, both because they may as a group be significantly different from the average, and their exclusion would bias the results, and because if "no information" cases are included for both independent variables, then the table will contain all cases for which the dependent variable was below its specified maximum.

It should also be noted that all values of the original variable are considered in defining the recoded variable. If a table uses a recoded variable, it does not matter whether the original variable exceeds its specified "maximum value" or not. Cases are included in a table so long as the value of the recoded variable is less than or equal to the maximum value specified for the recoded variable.

Each recoded variable is added to the original list of variables in sequence. Thus if the Format card identifies 5 variables, as in our example, the first recode variable becomes variable number 6.

The card set-up for card 6 is as follows:

Cols.	Parameter	Description
1-10	CATEGORY MAXIMA	The highest value from the original variable which is to be included in the first category of the recoded variable.
11-20	"	The highest value from the original variable which is to be included in the second category of the recoded variable. This category will include values higher than those in the first category, but no higher than the highest value listed for the second category.
21-30	"	etc.
31-40	"	etc.
.		One card 6 can identify 7 new categories
.		to appear in the recoded variable.
61-70		etc.
78-80		Sequence number.

If more than one variable is being recoded, first give the card 5 and card 6's for the first recoded variable. Follow the card 6 for the first recoded variable by the card 5 for the second recoded variable, followed by the card 6's for the latter. Continue similarly with succeeding recoded variables. If no recoded variables are to be created (so the value of RCDD on card 2 is 0), then omit cards 5 and 6.

Card 7 -- The Dependent Variable Card.

This program is designed to show the relationship between two independent variables and one dependent variable in any given table. As indicated in the introductory notes, the basic format of analysis is as follows: Two variables are cross-tablulated in such a way as to produce a matrix. These variables appear on the margins of the table and are referred to as the independent variables. The mean score on some designated variable for all respondents falling within a given cell is computed. The variable on which the mean is computed appears in the table cells, and will be referred to as the dependent variable. Thus, in the nine-cell table described in the introductory notes, Father's Education and Mother's Education are the independent variables, and the examination score is the dependent variable.

It is possible that the investigator may wish only to cross-tabulate two variables. This he would do by setting the parameter "dependent variable" to 0 in card 7. The table produced would be a frequency count (and percentages) of the row and column variable, that is, of the two "independent variables." Since percentages within each row sum to 100%, this table will show the effect of the row variable on the column variable.

Sometimes it is important in a study to analyze only part of the sample, or to compare the responses of two or more different groups within the sample. This program allows for this by use of a FILTER variable. A filter variable is a variable which the investigatory uses to sub-divide his sample. We sometimes say that he uses a control variable, or that he stratifies his sample. Presume that the investigator in the present study wishes to compare male and female students with respect to the relationship between family background and examination success. Variable 4, the sex of the student, would be used as a filter variable. This would allow the investigator to compute one set of tables for male students and another set for female students.

Card 7 identifies the dependent variable to appear in a table, and any filter variable which is to be used. It is followed by one or more card 8 which identifies the independent variables. The number of pairs of card 7 and 8 must agree with the number of dependent variables (DEPS) identified on card 1.

Card 7 first gives the sequence number of the dependent variable (in cols. 3-5). In our study, the examination score is the dependent variable and thus the number 5 would appear to identify the dependent variable. If a 0 is entered in cols. 3-5, this indicates no dependent variable and instructs the program to produce a frequency table using the row and column variables identified in card 8.

Card 7 next identifies the number of tables to be produced using the dependent variable. The number listed must take into account whether a filter variable is being used, which of course greatly increases the number of tables. For instance, if you simply wished to compute mean examination scores for all students classified by father's and mother's education, then you would produce only one table (which would be identified on card 8). If you wished to produce this table for both males and females, there would be two tables. If your filter variable had three categories, you would produce three tables, and so forth. Usually, the investigator wishes to make a series of tables on any given run. Assume that you wished to compare examinationsscores under varying conditions of family background. You might ask for the following tables:

1) examination scores by father's education and mother's education.
2) examination scores by father's education and number of educated siblings.
3) examination scores by mother's education and number of educated siblings.

You would be asking for three tables, and would need three card 8. If you were using sex as a filter variable, you would be asking for six tables. In the latter case, you would need two card 7's, each with a 3 in cols. 8-10 and each followed by three card 8's.

The total number of tables on card 7 must not exceed 100. Infact, the total sum of the number of tables on all card 7's must not exceed 100. As explained in Appendix 5, storage limitations often restrict this total to around 25.

Card 7 next identifies the sequence number of the filter variable to be used, a 0 is used if no filter variable is wished. Cols. 13-15 record the filter variable.

Card 7 next identifies the value of the filter variable to be applied to the following tables. (This appears in cols. 16-20 and is blank if filter = 0). If, for instance, you did wish a set of tables for male and for female students, then cols. 16-20 of the first card 7 would have a 0, instructing the program to create tables only for male students. Following the card 8 which went with this card 7, there would be another card 7, identical to the original one in every respect except that a 2 would appear in cols. 16-20, indicating that the next set of tables is to be created only for female students.

The set-up of card 7 is as follows.

Cols.	Parameter	Description
3-5	DEP	The sequence number of the dependent variable to be used for the following set of tables.
8-10	TABS	The number of tables to be used with this dependent variable, and the value of the filter variable specified below.
13-15	FLTR	The sequence number of the variable to be used as a filter. (0=no filter variable)
16-20	VALU	The value of the filter variable for cases to be included in the following tables card. (Blank if filter = 0)
78-80		Sequence number

Card 8 -- The Tables Card(s)

Card 8 identifies the independent variables. The total number of card 8s which follow any given card 7 must agree with TABS on that card 7. Card 8 first lists (in cols. 3-5) the sequence number of the row variable (the variable which appears on the side margin of the table) and then lists the sequence number of the column variable (in cols. 8-10), the column variable appears at the top of the table.

You can set either of these to 0, which would then produce the mean of the dependent variable by only one variable. For instance, if you simply wished to compare examination scores of male and female students, you would select variable 4 as your row or column variable and set the other to 0. (Which you used

is immaterial, the only difference is how the print-out is to be read; using the row variable would produce the table vertically and using the column variable would produce it horizontally.)

Perhaps it is now clearer why you set the maximum value (on card 4) to eliminate certain values, such as no information for instance. If the maximum value for the first two variables were set at 9, then the tables produced with father's and mother's education would include cells produced by the cross-tabulation of each "no information" code with all other codes. (Of course there may be times when you wish to have this information, and thus you would not eliminate the respondents punched 9 in variables 1 and 2). If the maximum were set at 9, there would also be rows and columns with values not used, i.e. 3 to 8. This would clutter the table, take extra computing time, and reduce the number of tables which could be produced.

The set-up for card 8 is as follows:

Cols.	Parameter	Description
3-5	ROW	The sequence number of the independent variable to be tabulated in the row of the table using the dependent variable listed above. (0=no tabulation by rows.)
8-10	COL	The sequence number of the independent variable to be tabulated in the column of the table using the dependent variable listed baove. (0=no tabulation by columns.)
78-80		Sequence number.

Each card 8 produces a separate table. All card 8s which follow a specific card 7 produce tables using the dependent variable there identified. If card 7 also identifies a filter variable, then the card 8s will produce tables only for that part of the sample which has the value specified by the filter variable parameter, VALU.

A series of card 7's and card 8's can be used. Each card 7 changes either the dependent variable or the filter variable. (Or of course it may change both.)
The total number of tables must not exceed 100.

Card 9 -- Data Cards (optional)

If your data are in card format (UNIT = 1 on parameter card),

then place the data deck at the end of the last card 8. The number of cases must agree with the number appearing in cols. 1-5 of the parameter card. (The number of cases of course is not the same as the number of cards, for each case may take more than one card.)

If data are not on cards, the name of the magnetic tape or paper tape to be mounted must be given to the computer operator when submitting the "job".

Control cards 1 through 8 may appear as many times as required exceed any of the limitations on parameters or storage (see Appendix 5). The excess tables would be produced by a second set of analysis cards. A new complete set of control cards is required, beginning with Card 1. No parameters, formats, or data except for information about the magnetic data tape are saved. If a magnetic tape was used, it will be rewound for the second (or third, etc.) set of analyses according to the value of REW on the parameter card for the first set of control cards. Only one magnetic tape can be used for any given series of runs. If the data are recorded on cards or paper tape, then they must be read in a new.

Card 10 -- The Termination Card.

This card is placed at the end of the last set of analyses and terminates the execution of the program. The word FINISH is typed into the first six columns.

The set-up of card 10 is as follows:

Columns	Description
1-6	The word FINISH
78-80	Sequence number.

Appendix 1: Example on Tourist Expenditure.

A survey of tourist expenditures has been conducted in order to estimate average expenditure and length of visit for various groups of tourists visiting Kenya. The original survey data and some derived variables have been put on a magnetic tape entitled TOURST SRVEY.

The relevant input variables are as follows:

Variable number	Columns	Description	Maximum
1	4-8	Month and identification number.	12999
2	9-10	Nationality	5
3	11-12	Purpose	5
v 4	13-17	Length of visit (days)	150
5	18-22	Nights in Hotel	150
6	30-32	Persons in party	15
7	33-34	E A Resident	1
8	35-41	Expenditure per visitor day (Shs)	1000
9	42-48	Proportion of nights in hotel	1

Although some variables have decimal parts, they will automatically be read correctly because the program reads the decimal point.

The Format statement for these variables is:

(3X,F5.0,2F2.0,2F5.0, 7X,F3.0,F2.0,2F7.0)

Some of these variables require recoding for the following reasons: (1) so that a row or column is not assigned to the value "0" in original variables 2 and 3; (2) to combine values 4 and 5 of variable 3; (3) to divide variables 4,5, and 9 into a few meaningful categories. The recoding parameters are as given below. The open-ended category has been used for variables 12 and 13, as described on page 10.

Variable Number	Original Variable	Max. Value	CATS (Categories)	Category 0	1	2	Maxima 3	4
10	2	4	5	1	2	3	4	5
11	3	3	4	1	2	3	5	
12*	4	5	5	0.5	7	14	21	28
13*	6	2	2	1	2			
14	9	2	3	0.2	0.8	1.0		

The following tables are needed:

(1) With expenditure (var 8) as dependent variable, tabulated by
   a) nationality, purpose, and residence simultaneously.
   b) nationality, proportion in hotel, and residence
   c) length, persons, and residence.

(2) With length (var 4) as dependent variable tabulated by
   a) nationality, purpose, and residence.

(3) With nights in hotel (var 5) as dependent tabulated by
   a) nationality and purpose.

Since the tabulations in (1) and (2) are 3-way, one of the variables must be selected as filter. The computation will proceed fastest if the variable with the fewest categories (i.e. residence) is named as filter. Thus the tabulations in (1) require separate dependent variables. The control cards as originally punched are shown below. Following is a listing by the program S04B of these cards with the program's interpretation of the parameters.

Output of S04B

Card
type

```
1  S04C - STUDY OF TOURIST EXPENDITURE.  SHEPARD/MARSHALL
   CASES VARS DEPS UNIT RCOD LIST REW TVAR              TAPE
2    958    9    7    2    5    1   0   14      TOURST SRVEY
3  (3X,F5.0,2F2.0,2F5.0,7X,F3.0,F2.0,2F7.0)
3
   MAX VALUES
4  12999    5    5  150  150   15    1 1000  1   4    3    5
       2    2

   (The max values of the last two variables have been moved due
        to space limitations)
   RCOD VAR (VAR NO 10)
   ORIG CATS
5      2    5
        CATEGORY MAXIMA
6           1.        2.        3.        4.        5.
   CODE        RANGE (VAR NO  2)
      0 -99999.99 TO       1.00
      1      1.01 TO       2.00
      2      2.01 TO       3.00
      3      3.01 TO       4.00
      4      4.01 TO       5.00
```

(Following the recoding instructions for a recoded variable, the program indicates the values of the original variable included in each category of the recoded variable, as shown above. The print out of cards 5 and 6 for the other recoded variables has not been reproduced here.)

```
      DEP TABS FLTR VALU
7       8    3    7    0
      ROW--COL
8      10   11
8      10   14
8      12   13
```

(The card 7 shows three tables are to be made with variable 8 as dependent variable, and including cases with variable 7 having a value of 0. Thus these tables will show average expenditure for non residents of East Africa. The card 8's which follow it give the variables to be tabulated in the rows and columns; nationality (after recoding) is tabulated in the rows of the first two tables. The variables purpose, fraction of nights in hotel, and persons are tabulated in columns of successive tables. The variable under "ROW" acts as a control variable, since percentages will be based on row totals.)

```
7     DEP TABS FLTR VALU
        4    1    0    0
      ROW--COL
8      10   11
```

(On this second card 7, the 0 in the FLTR field indicates the table is to be made without a filter. The 0 beneath VALU is meaningless. The groups of cards for the five other dependent variables have not been reproduced here. After reading the last card 8, the program began reading the data, which for this job was on magnetic tape. Since the parameter LIST was set to 1, the program listed cases in which some variable exceeded its specified maximum. Of the 15 cases which were listed, two are reproduced and discussed below.)

```
D-RANGE, C=   72 V=   8 1320 1002
*   72 *
 7093.   .5.   5.   .1.   1.   2.   1. 1320.   1.   4.
   3.   1.   1.   2.
```

(Due to space limitations, the values of variables 10 to 14 are shown below rather than to the right, of variables 1 to 10. The case shown above is the seventy - second case. The range check was activated by variable 8 with a value of 1320. The number 1002 is the maximum value of variable 8, plus 2. (See Appendix 6). The advantage of including the month and identification number as variable 1 is now apparent. This suspicious case showing very high expenditure (shs 1320 per visitor day) can now be checked, It is month 7, tourist 093. The variables 10-14 are recoded variables. By referring to the original variables, one can verify that the recoding is correct.)

```
D-RANGE, C=  792 V=    9    8    3
D-RANGE, C=  792 V=   14    3    4
*  792 *
 9156.    1.    5.    1.    8.    2.    0.  814.    8.    0.
    3     1.    1.    3.
```

(This is case 792, month 9, tourist 156, according to variable 1. Both variables 9 and 14 had range errors. The fraction of nights in hotels is 8, because the tourist indicated that he stayed 8 nights in hotels but 1 night total. This is probably an error in coding which S04B has detected. Finally the computer prints the tables. Below one table is reproduced and discussed.)

TABLE NO 2: VAR NO 10 (ROWS), VS VAR NO 14 (COLS). MEANS OF VAR NO 8

FILTER--INCLUDES ONLY CASES WITH VAR NO 7 = 0

(Since variable 7 is "Resident of EA" this table will include only visitors who indicated they were not residents of East Africa. Before analysing the table, the user should label the rows and columns by referring to the recoding categories and his original coding instructions as has been done below.)

Nationality		Proportion of nights in hotel: 0 to 20% (0)	21% to 80% (1)	81% to 100% (2)	Total
U K	0 *	47.38	85.61	133.21	95.07
	(N)	84	16	109	209
	(%)	40.2%	7.7%	52.2%	100.0%
Other Europe.	1 *	76.50	93.94	156.21	136.68
	(N)	33	7	117	157
	(%)	21.0%	4.5%	74.5%	100.0%
USA	2 *	82.49	210.45	187.92	174.08
	(N)	43	12	253	308
	(%)	14.0%	3.9%	82.1%	100.0%
Tanzania/ Uganda	3 *	1.00	0.00	0.00	1.00
	(N)	2	0	0	2
	(%)	100.0%	0.0%	0.0%	100.0%
Other	4 *	50.28	58.87	191.32	139.77
	(N)	52	3	95	150
	(%)	34.7%	2.0%	63.3%	100.0%
Total	999 *	59.20	124.46	171.63	140.33
	(N)	214	38	574	826
	(%)	25.9%	4.6%	69.5%	100.0%

(In each cell is shown the mean, the frequency, and the frequency as a percent of the row total. As indicated in the table heading, the mean will be of variable 8, expenditure per visitor day. The upper left cell shows visitors who are UK citizens, non residents of East Africa (because of filter) who spent 20% or less nights in hotels. These visitors spent an average of shs 47.38 per visitor day; they numbered 84, representing 40.2% of the UK citizens in the table. The computer then prints the variance of dependent variable within each cell, shown below.)

TABLE OF CELL VARIANCES

	0	1	2	
0 *	5300.	3084.	14783.	11701.
1 *	17331.	4794.	30274.	27385.
2 *	25728.	73751.	32928.	34576.
3 *	0.	0.	0.	0.
4 *	6881.	675.	41123.	32928.
999 *	11652.	27566.	30654.	27936.

FOR ALL CASES IN TABLE N= 826, MEAN= 140.330, SD= 167.140

(The main use of this table is to assist in interpreting the means given above. Notice that the variance* in row 2, column 1, 73751, is much higher than any other variance. This indicates that in this cell expenditure was particularly dispersed. The relatively high mean for this cell, 210.45, is probably due to the fact that there were a few visitors in this cell with very high expenditures Thus the cell variances enable the investigator to decide whether a cell mean is representative, or due to a few extreme observations. The variance may also be used statistically in a t-test or an F-test to determine whether two cell means are significantly different.)

The following conclusions can be drawn from the table of means above: (1) Visitors from outside EA are ranked in descending order of expenditure as follows: USA, other, other European, UK. (The sample of Tanzania and Uganda citizens is too small for interpretation). (2) Visitors spend either very few (under 21%) or very many (over 80%) of their nights in hotels. (3) The proportion of nights in hotels seems to vary with nationality in almost the same way as expenditure. (4) The nationality group "other" is heterogeneous: those visitors staying outside hotels spend little; those in hotels for most nights have the highest expenditure rate of any group. (5) Among visitors staying the same proportion of nights in hotels, Americans are still highest, followed by other Europeans, and British. (6) The sample contained 826 visitors. Of the 958 total cases, the balance were excluded from this table because they were residents of E A or the residence code was invalid, the proportion of nights in hotel was invalid or expenditure exceeded shs 1000/-. Of those visitors in the table the mean was shs 140/33 and standard deviation (the square root of variance) 167/14. If expenditure is normally distributed, then 5% of these visitors must exceed shs 400/- per day. (7) Extrapolating from the "marginal" means on the column variable, one can conclude that a non-resident of E A spends shs 50/- per day staying outside a hotel (i.e. in a private home or campground) and shs 125/= more if he stays in a lorge or hotel.

---

* Cell variance is defined as follows:

Let y_i be the value of the dependent variable for the $i^{\underline{th}}$ case in the cell.

Let $\bar{y}$ be the mean of y_i in the cell.
Let N = number of cases in the cell.
Cell Variance = $\frac{1}{N-1}\left(\sum_{i=1}^{N} y_i^2 - N\bar{y}^2\right)$.

Appendix 2: Program Structure

The structure of S04B is as follows: Each "run"or "submission" of the program may perform any number of sets of analyses. Each set is a group of recodings and tables which can exhaust the full capacity of the program. After making all the tables requested as one set, the program clears its memory and goes back to the beginning accepting the new complete set of control cards as if it were a new submission.

Within a set, the user specifies maximum values and recoding instructions, which applies to the whole set. Then he specifies the first dependent variable; the filter, if any, applies to all tables with that dependent variable. Finally, he specifies the row and column variables for tables with that dependent variable and filter. He then specifies the second dependent variable, its filter, and the tables to be made with it, next he specifies the third dependent variable, etc.

Schematically, this may be illustrated as follows:

First set of analyses - "Set 1"
Parameters, Format, Max values,
Recoding (optional)

First dependent variable
Filter (optional)
Table 1
Table 2

Second dependent variable
Filter (optional)
Table 3
Table 4
Table 5

Third dependent variable
Filter (optional)
Table 6
Table 7

Second set of analyses - "Set 2"
Parameters, Format, Maximum values
Recoding (optional)

First dependent variable
Filter (optional)
Table 1
Table 2
Table 3
Second dependent variable etc
etc.
Third set of analyses - "Set 3"
etc
Termination card.

Appendix 3: Control Card Formats

General comments: All numbers must be right-justified. In general, five columns have been allowed for numbers. Numbers may not be punched with a decimal point except on card 6 (Category Maxima), where the decimal point is optional and ten columns are allowed for each number.

Columns 73-80 of all control cards are reserved for sequence numbers. The first card should be labeled 010 in cols. 78-80, the second 020, the third 030, etc. This enables control cards to be identified easily.

1. Title card

Cols.	Description
1-4	Name of job (e.g. S05Z)
16-65	Description of job (i.e. user's name, project, description of set of analyses)
78-80	Sequence number

2. Parameters card

Cols.	Parameters	Description
1-5	CASES	Number of cases. ($1 \leq$ CASES ≤ 99999)
8-10	VARS	Number of input variables (independent and dependent) ($1 \leq$ VARS ≤ 100)
14-15	DEPS	Number of dependent variables (cards). ($1 \leq$ DEPS ≤ 20)
20	UNIT	Input unit for data 1 = cards; 2 = magnetic tape; 4 = paper tape.
24-25	RCOD	Number of variables to be created by recoding input variables ($0 \leq$ RCOD ≤ 40.)
30	LIST	List data cases? 0 = No list; 1 = list all cases for which any variable exceeds its specified maximum value; 2 = list all cases.
35	REW	Rewind magnetic data tape at the end of this set of analyses (after card 8)? 0 = no; 1 = yes.
38-40	TVAR	Total number of variables (VARS + RCOD). ($1 \leq$ TVAR ≤ 100)
49-60	TAPE	Name of magnetic data tape. (Blank if UNIT not 2)
78-80		Sequence number (i.e. 020)

3. Format Cards (two cards required)

Columns	Description
Card 1:	
1-72	Format statement, as described in Appendix 4. number of "F" fields must agree with VARS.
78-80	Sequence number
Card 2:	
1-72	Continuation of format statement if required. Otherwise blank.
78-80	Sequence number

4. Maximum Values Card(s)

Columns	Parameter	Description
1-5	MAX VALUES	The maximum value of the first variable
6-10	" "	The maximum value of the second variable
.		
.		
.		
66-70	" "	The maximum value of the fourteenth variable.
78-80		Sequence number.

Notes: -1= Maximum unknown or unspecified. (Such variables may not be used as independent variables, Card 8.) Use additional cards as necessary. Number of MAX VALUES must agree with TVAR.

5. Recoded variables Card.

Note: Omit if RCOD = 0. For each recoded variable, use one Card 5, followed by one or more Card 6. The total number of Card 5's must agree with RCOD.

Cols.	Parameter	Description
3-5	ORIG	Sequence number of original input variable to be recoded. ($1 \leq ORIG \leq VARS$)
6-10	CATS	Number of categories created on the recoded variable.
78-80		Sequence number

6. Category Maxima. Card(s).

Note: The number of CATEGORY MAXIMA must agree with CATS on the preceding Card 5. Use additional cards if necessary, all with the layout below:

Cols.	Parameter	Description
1-10	CATEGORY MAXIMA	The maximum value of the original variable to be included in the first category of the recoded variable.
11-20	"	The maximum value of the original variable to be included in the second category of the recoded variable.
21-30		etc.
31-40		etc.
⋮		
61-70		etc.
78-80		Sequence number

7. Dependent Variable Card.

Note: For each dependent variable, use one Card 7, followed directly by one or more Card 8's. The number of groups of Card 7 and 8 must agree with DEPS of Card 1.

The parameters on this card apply to all of the TABS tables specified on the Card 8s following. A separate dependent variables card is required for each set of tables having a different dependent variable, filter variable, or filter value.

Cols.	Parameter	Description
3-5	DEP	Sequence number of the dependent variable. (0 = no dependent variable. Only frequencies and per centages tabulated.)
8-10	TABS	Number of tables to be produced. ($1 \leq$ TABS ≤ 100. Also total of TABS on all Card 7's ≤ 100.)
13-15	FLTR	Sequence number of variable to be used as the filter. (0 = no filter variable.)
16-20	VALU	Value of filter variable for cases to be included.

8. Tables Card(s)

Note: Number of Card 8's must agree with TABS of preceeding Card 7.)

Cols.	Parameter	Description
3-5	ROW	Sequence number of independent variable to be tabulated in rows of table, with dependent variable above. (0 = no tabulation by rows. $0 \leq$ ROW $\leq$ TVAR.)

8-10	COL	Sequence number of independent variable to be tabulated in columns of table with dependent variable above. (0 = no tabulation by columns. $0 \leq COL \leq TVAR$.)
78-80		Sequence number

9. Data Cards.

(Optional.) Insert data cards here if UNIT = 1. Otherwise name of magnetic or paper tape to be mounted must be given to operator. Data cards must be in the format indicated by the Format Card (Card 3), and the number of cases must agree with CASES on Card 1.

1 to 9 New set of control cards

(Optional, as many times as required.)

If the tables required would exceed any of the limitations on parameters or storage, then those tables must be produced as an additional set. A new complete set of control cards is required, beginning with card 1. No parameters, formats, or data except for information about the magnetic data tape, are saved. If a magnetic data tape was used on a previous set, then a different tape cannot be used on this set. The tape will be rewound or not, according to the value of REW in the previous set. Data on cards or paper tape, if any, must be read in anew..

10 Termination card

Cols.	Contents
1 - 6	The word 'FINISH'
78 - 80	Sequance number

This card placed at the end of the last set of analyses, terminates execution of the program.

Appendix 4 - Format Statements

The Format statement, of the same type as in FORTRAN programs, indicates the columns for each variable. To write a Format statement for S04B, begin with an opening parenthesis on the first Format Card. Use the letter "X" for "skip columns", the letter "F" for "read a variable", a slash "/" for "go to the next card for more data in the same case", and a comma "," for separating fields within the same card. Precede an "X" by the number of columns to be skipped; follow an "F" by the number of columns for the variable, then by ".0". You may precede the "F" with a number to indicate the number of consecutive variables of the width specified which are to be read. End the Format statement with a closing parenthesis.

For example, the Format Statement

(6X,F1.0,F2.0/10X,10F1.0,20X,F5.0)

Means: Skip six columns, read a variable of one column, and read a variable of two columns; go to the next card, skip ten columns, read ten variables each of one column, skip twenty columns, read a variable of five columns.

The program S04B cannot process alphabetic data. Columns containing alphabetic data must be skipped. Two cards are <u>always</u> required for the format statement in S04B. The second format card is treated as a continuation of the first, with column 1 of the second card treated as if it were col.73 of the first. If the Format statement in completed on the first card, then the second should be blank in columns 1 to 72.

Appendix 5 - Storage Limitation.

In order to run quickly and efficiently on the computer, this program has been written to create all the tables specified in a single set of analyses in one pass through the data, without using overlays or magnetic tape backing store. As a result, there is a fairly severe storage limitation on the number and size of tables created in one set. This limitation is in addition to the restrictions on the parameters in card 2.

A total of 2400 computer storage locations (words) are allocated for the combined storage of category maxima for recoding variables, as well as table creation. The maximum number of tables which can be created as one pass depends on the size of the table. It is important that the user not try to exceed this maximum, for the program will detect the error in the input phase, and halt before creating any tables.

The table below shows the maximum number of tables of certain sizes which can be made in one set. For table dimensions not given, or for sets with substantial recoding (more than 9 recoded variables), consult the formulae below instead.

Maximum number of Tables in One set

Dimensions of Table*	Number of Tables
5 x 5	31
5 x 7	22
7 x 7	16
7 x 9	12
9 x 9	9

Example: We wish to make a number of filtered tables of examination scores cross tabulated by father's and mother's education. Suppose, as in the program exposition, the maximum values of these variables are 2 and 2. The dimensions of each table are then 4 x 4. The next largest dimensions given above is 5 x 5, for which 31 tables could be made in one set. Thus at least 31 4 x 4 tables could be made in oneset.

---

* The dimensions of a table are the total number of rows and the total number of columns. Since a column is alloted for each value of the column variable from zero to its maximum value, inclusive, and to totals, the column dimension equals the maximum value of the column variable + 2. Similarly, the row dimension equals the maximum of the row variable + 2.

The storage requirement may be calculated precisely with the following formulae:

1. Storage for CATEGORY MAXIMA:

$$\sum_{\substack{\text{For all}\\ \text{recoded}\\ \text{variables.}}} \text{CATS}$$

2. Storage for tables with mean of dependent variable:

$$3 \times \sum_{\text{all tables}} \left[(\text{MAX VALUES of var ROWS}) + 2 \times \right] \left[(\text{MAX VALUES of var COL}) + 2 \right]$$

3. Storage for tables with frequency counts only (dependent var 0):

$$\sum_{\text{all tables}} \left[- (\text{MAX VALUES of var ROW}) + 2 \right] \left[\text{MAX VALUES of var COL} + 2 \right]$$

4. The sum of items 1, 2, and 3 above must be $\leq$ 2400.

Appendix 6 - Error Messages and Data Listing.

If any of the values on the control cards 1 to 8 are outside their valid ranges, or any program or storage limitations are exceeded, the program will print an error message after reading and printing the offending card. After an error, the computer terminates execution of the program. The last line printed by the program will be as follows:

ERROR STOP N X Y

N is the error number, and X and Y are both numbers, all explained in the table below. This error stop indicates either that the value of X is negative or zero, or that it exceeds its permitted maximum Y. The user should determine the cause of the error and correct it before resubmitting his job.

N	X	Y	Remarks; likely cause
1	CASES	99999	CASES zero or negative. Cards out of order
2	VARS	100	More than 100, or zero, input variables
3	DEPS	20	More than 20 or zero dependent variables.
4	UNIT	4	Invalid unit number
5	RCOD + 1	41	More than 40 recoded variables.
6	LIST + 1	4	Invalid value for LIST.
7	REW + 1	2	Invalid value for REW.
8	TVAR	100	More than 100, or zero, total varaiables.

N	X	Y	Remarks
9	Number of current table	100	Too many tables requested.
10	Not used		
11	DEP + 1	TVAR + 1	Invalid dependent variable sequence number.
12	ORIG + 1	TVAR + 1	Invalid original variable sequence number.
13	FLTR + 1	TVAR + 1	Invalid filter variable sequence number
1001 to 1100*	ROW + 1 or COL + 1	TVAR + 1	Invalid variable sequence number on Table Card.
2001 to 2100*	Storage cells used	2400	Too many tables requested in one set of analyses. Create separate sets. (see Appendix 5.)

Another possible error is the following:

EXECUTION ERROR 0.

This indicates an error while attempting to read the data. Either the Format Statement is not valid, or the program has encountered an alphabetic character in one of the columns of the input variables. Examine the data deck carefully to determine whether the computer operator noted the last card read. Otherwise, check data and format statements or re-run with LIST = 2.

If list is 1 or 2, the program will also list range "errors" in the data. These cause the case to be excluded from tables using the variable with the range error. It will be included normally in other tables, however. A range error in an original variable does not effect tables based on a derived recoded variable, unless that recoded variable has a range error as well. The computer will not stop due to range errors.

The error message printed on data range checking is given below. If required, range errors may be noted on several variables within a single case.

D-RANGE, C = a V = b

where: a = number of case with range error
b = sequence number of variable with range error
c = value of variable b
d = 2 + specified maximum value of variable b (both input and recoded)

* The last three digits of N indicate which tables card, counting from the first, caused the error.

This is followed by the values of all variables (both input and recoded) for the case with the range error, in the form below. If the value of LIST = 2, then the print-out below appears for every case regardless of whether it had a range error.

Case number, enclosed in asterisks.
Values of variables 1 to 20 in order.
Values of variables 21 to 40, in order*
" " " 41 to 60, " " *
" " " 61 to 80 " " *
" " " 81 to 100 " " *

---

* Omitted if not required by number of variables.

Appendix 7 - Special Facilities of S04B

The program has the flexibility to perform special types of tabulations which at first might not occur to the user.

1. Excluding Certain Cases

In some cases you may wish to use the filter option in order to exclude a certain set of cases. This can be done as long as the cases to be excluded have either the lowest or highest coded number(s). A recoded variable (RCOD) can then be created which has two values (0,1), one of which contains all the cases to be excluded and the other, all those to be included. The following three examples are illustrative. Imagine that the data is coded from 0 to 9. (a) In order to exclude all cases punched 0, one would create a recoded variable with the maximum in the first category of 0 and in the second category of 9. The filter statement would then specify the inclusion of only those cases punched 1 on the recoded variable. (b) In order to exclude the cases punched 8 and 9, one would set the category maxima at 7 and 9 and specify 0 as the inclusive filter value. (c) In order to exclude the cases punched 0 and 9, one would set the category maxima at 0 and 8 and specify 1 as the inclusive filter value.

2. Analyzing the sample in sub-groups.

Suppose one wanted a series of tables for each of two sub-groups, say males and females. It would be more convenient to sort the data so that the two sub-groups were separate, say all males, followed by all females. This sorting would be accomplished prior to analysis using S04B.

The required tables would be run as two sets of analyses. First tables on the males would be made, for which CASES is the number of male cases, followed by tables for females, in which CASES is the number of female cases. If the data are on a tape, then the rewind parameter is set not to rewind the tape after the first set (i.e. REW is 0). In this way, the computer will be able to start from inside the tape reading the female data on the second set.

Dividing the sample in this way frees the filter facility for some other use and allows more types of tables to be made in each set.

Appendix 8 - Computation Time

The process which takes the most time is reading the data from cards or tape. Since this must be done anew with each set of analyses, the number of sets should be as small as possible. The table below provides a rough indication of running times from experience on the University of Nairobi's. ICL 1902A computer. The Nairobi computer has floating point hardware.

Execution Times for S04B

Process	Computing time (minutes)
1. Set up, loading the program and tapes, reading control cards.	2
2. Reading and processing data per 100 cases	
a) for set with 10 tables	1
b) for set with 25 tables	1½
3. Preparing and printing tables per 6 tables	1

Appendix 9: Source Program Description

The source program S04B has been written in Fortran IV language for the ICL 1900 series. It uses a few expressions which are not standard Fortran i.e. subscripts may be themselves contain subscripts or complex mathematical statements, and the parameters of a "DO" statement may be expressions, instead of single variables.

The program has been constructed with a master segment, called S04B, and five subroutines. The master segment reads the title card and calls subroutines PARAMS, RECORD, and TABLE. Subroutine PARAMS reads and processes the program control cards; RECORD reads and processes the data records, and TABLE creates and prints each table. The small subroutines CHECK and VF do range checks on parameters and check for division by zero, respectively.

The user wishing to delve more deeply into the source program may find the following information on the storage of parameters and data helpful.

Arrays in S04B

Array (dimensions)	contents
N (I) (8)	parameters I = 1 CASES I = 5 RCOD 2 VARS 6 LIST 3 DEPS 7 REW 4 UNIT 8 TVAR
M (I) (10)	Parameter maxima I=1,8 Maximum value of parameters in N 9 Maximum number of tables (100) 10 Maximum storage cells (2400)
TITLE (I) (9)	Title card (alphabetic format)
SAYOUT (I) (18)	Format cards (alphabetic format)
MV (K) (101)	Dimension of variable K* = maximum value (variable K) + 2.
MR (I,IR) (2,20)	Recoded Variable Matrix I = 1 ORIG 2 Subscript in T of last category maximum IR = Recoded variable counter
MD (I,ID) (4,20)	Dependent variables matrix I = 1 DEP 2 Last table in MT with this dependent variable 3 FLTR 4 VALU ID = Dependent variable counter
MT (I,IT) (4,100)	Tables matrix I = 1 ROW 2 COL 3 Dimension of row variable, MV (ROW) 4 Origin - 1 in T array of beginning of table IT = table counter

* Variable 0 transformed to variable N(8)+ 1 with dimension 0.

TAPE (I) (2)	Name of tape.
V (K) (100)	Values of variables in current record (floating point)
IV (K)	Values of variables, fixed point.
T (I) (2400)	Dynamic storage array for category maxima and tables 1. First elements: category maxima for recoded variables, referenced by MR (2,IR) 2. Later elements: tables, referenced by MT (4,IT) and called TAB.
TAB (I,J,K) (within T)	Semi-compiled table I = 1 frequency for cell, $\sum N$ 2 sum for cell, $\sum Y$ 3 sum of squares for cell, $\sum Y^2$ J = row of cell K = column of cell.